UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF TEXAS
WACO DIVISION

PARUS HOLDINGS INC., Plaintiff, v. APPLE INC., Defendant.	C.A. No.: 6:19-cv-00432-ADA **(Lead Case)** JURY TRIAL DEMANDED
PARUS HOLDINGS INC., Plaintiff, v. GOOGLE LLC, Defendant.	C.A. No.: 6:19-cv-00433-ADA (Consolidated Case) JURY TRIAL DEMANDED
PARUS HOLDINGS INC., Plaintiff, v. LG ELECTRONICS INC. and LG ELECTRONICS U.S.A., INC., Defendants.	C.A. No.: 6:19-cv-00437-ADA (Consolidated Case) JURY TRIAL DEMANDED

PARUS HOLDINGS INC., Plaintiff, v. SAMSUNG ELECTRONICS CO., LTD. and SAMSUNG ELECTRONICS AMERICA, Defendants.	C.A. No.: 6:19-cv-00438-ADA (Consolidated Case) JURY TRIAL DEMANDED

PARUS HOLDINGS INC., Plaintiff, v. AMAZON.COM, INC., Defendant.	C.A. No.: 6:19-cv-00454-ADA (Consolidated Case) JURY TRIAL DEMANDED

DECLARATION OF BENEDICT OCCHIOGROSSO IN SUPPORT OF PLAINTIFF PARUS HOLDINGS INC.'S OPENING BRIEF REGARDING CLAIM CONSTRUCTION

I, Benedict Occhiogrosso, declare as follows:

1. My name is Benedict Occhiogrosso. I am a citizen of the United States over 18 years of age, and the co-founder and President of DVI Communications, Inc. I make this declaration in support of the Opening Brief Regarding Claim Construction of the Plaintiff in the above-captioned matter, Parus Holdings, Inc. ("Parus" or the "Plaintiff"). I have personal knowledge of the facts set forth below, and, if called upon as a witness, I could and would testify to such facts under oath.

2. I have been retained by counsel for Parus as an expert in this consolidated case to assess the proper claim construction of the disputed terms of United States Patent No. 7,076,431 (the

"'431 patent") and United States Patent No. 9,451,084 (the "'084 patent") (together, the "Asserted Patents"). As I explain in greater detail below, I believe that I am at least one of skill in the relevant technology and art of the Asserted Patents, and I offer my analysis and opinions regarding my understanding of the meaning of the disputed claim terms of the Asserted Patents as such a person skilled in the art.

3. I am being compensated by Parus at my hourly rate of $475. My compensation does not depend in any way on the outcome of this case. Nor does it depend in any way on the outcome of any one or more disputed claim terms, or the particular opinions I express, or the testimony I give.

4. This declaration contains my conclusions, a summary of my analysis, an overview of my qualifications as one skilled in the art of the Asserted Patents, an overview of the materials I have considered in arriving at my conclusions, an overview of the terminology and legal principles that I applied in my analysis, an overview of the technical background of the subject matter, and an overview of the Asserted Patents. It also includes my opinions regarding how one skilled in the art would understand the disputed claim terms in view of the claims, the teaching of the common specification, the patents' respective file histories, and, where appropriate, certain extrinsic evidence.

5. This declaration is based on information currently available to me. I reserve the right to offer additional analysis and opinions once I have had the opportunity to review Defendants' positions once they are disclosed in briefing that has not yet been exchanged among the parties. I also reserve the right to respond to the expert opinions (that are to date undisclosed) offered by Defendants in support of their claim construction positions, if any. I also reserve the right to include a review of additional extrinsic documents and information that Defendants may yet produce. Therefore, I expressly reserve the right to expand, modify, or supplement my opinions in view of and in response to any additional information that becomes available to me, any matters raised by Defendants, and/or

other opinions provided by Defendants' experts, or in light of any relevant orders from the Court or other authoritative body.

I. Summary of Opinions

6. As discussed in more detail below, I offer what I believe is the understanding one having ordinary skill in the art would have for the disputed claim terms offered for construction in this case.

7. In addition, in my opinion, Parus's proposals for the disputed claim terms best comport with the understanding a person of ordinary skill in the art would have had at the time of the invention. In particular, I agree with Parus's proposed articulations for the terms: (a) "voice enabled device," (b) "speaker-independent speech recognition device," and (c) "recognition grammar." In my opinion, Parus's proposal for each claim term best aligns with the understanding of the meaning of each such claim term by one skilled in the relevant art and the time of invention, in the context of the claims and the teachings of the specification. By contrast, I further opine regarding why I believe that Defendants' inconsistent proposals are inadequate, incorrect, and/or do not fairly capture the scope of the claim terms as drafted.

8. As for Defendants' proposed claim terms for construction, it is my opinion that no particularized construction need be given for certain of the proposed constructions including, for example, the term "web site," because the plain language of such terms would be readily understood by one skilled in the art at the time of invention. In my view, the plain and ordinary meaning of these terms should apply, and Defendants' proposed constructions for such terms diverge from the plain and ordinary meaning without justification, are artificially narrow, and do not adequately capture the proper scope of the patent terms as claimed.

II. Background and Qualifications

9. I offer the following summary description of my background and experience, which I believe qualifies me to opine as one skilled in the art of the relevant technology. I attach my CV as **Exhibit A** to this Declaration, which is current as of this date to the best of my knowledge.

10. I hold a Bachelor of Science Degree in Electrical Engineering as well as a Master of Science Degree in Electrical Engineering, both from the Polytechnic Institute of Brooklyn (now part of New York University).

11. I have authored or co-authored nearly three dozen articles in peer-reviewed journals, conference proceedings, texts, industry trade publications, and monographs. These publications span a range of topics including: Integrated Voice–Data Communications/Switching, Integrated Packet-Circuit Switching, Voice Digitization, Packet Voice, Indoor Wireless distribution, Disaster Recovery and Business Continuity, Data Center Engineering, Switching Processor Architecture, Telephone and Voice Mail Systems, PBX & LAN switching premises-based systems and related technologies and Internet of Things (IoT).

12. I have more than 40 years of telecommunications and information technology experience. I am the co-founder and President of DVI Communications Inc., a telecommunications and information technology and business consulting firm. Since establishing DVI in 1979, I have planned, designed, implemented, and managed large-scale projects involving wired and wireless communications systems, which included transmission of voice and data. Prior to founding DVI and for several years thereafter, I held a Department of Defense security clearance and worked on several classified programs within the defense industry, where I supported the development of several pioneering technologies that have served as the prototypes for many telecommunications and IT systems later utilized in commercial practice.

13. Throughout the course of my professional career, I have worked and consulted extensively in the fields of information processing, including transaction processing systems design and development (including centralized and distributed, encompassing both fault-tolerant and real-time processing), various hardware and software platforms (including enterprise client/server incorporating different OS & DBMS), systems integration, and telecommunications. These telecommunications systems include wide area and local area networking technology, data, voice, and video switching systems, and both wired transmission (such as cable and fiber optic) and wireless transmission (such as UHF/VHF, Microwave, satellite, cellular, and Wi-Fi, for private networks, Intranets, and the Internet).

14. My technical expertise encompasses several disciplines, such as transaction processing systems design for various industries and applications, IT platform engineering, data center, and telecommunications network design. In addition, I have extensive experience with numerous interfaces and protocols, real-time synchronization as well as selected applications systems in various industries. These include financial services (including call centers, trading floors, supporting consumer, commercial and investment banking), health information systems (including patient information, clinical and laboratory systems), transportation and energy utilities (including dispatch, customer service, and trouble reporting), military aerospace (including Command and Control, position tracking and fusion centers), and telecommunications switching and operational support systems (order entry, billing, trouble reporting). Many of these solutions have included voice responsive systems. I have also provided operational and strategic planning consultation for client IT enterprise-level systems. Over the course of my career, I developed hardware and software, and I have also managed engineering development teams engaged in product development.

15. I have extensive expertise in voice-data-video switching, and transmission systems deployed in networks, including both circuit switching and packet switching using wireline and wireless distribution methods (including Land Mobile radio, Satellite, microwave, cellular, and Wi-Fi). In addition, I have developed various applications systems including voicemail, e-mail, unified messaging, and audio/video recording for a variety of facility types including call-contact centers, data centers, trading floors, and mission-critical communications centers. I have detailed knowledge of Internet Protocol (IP) technology in general, and also Voice over Internet Protocol (VoIP). I have extensive expertise in voice digitization, buffering, relay and delay techniques, as well as digital signal processing of voice signals. At present, my primary responsibilities encompass strategic planning and systems design of client IT Infrastructures and program management for major projects undertaken by DVI.

16. I have also developed technical expertise with regard to telecommunication applications utilizing speech recognition, including call center and interactive voice response applications, like the kind described in the Asserted Patents. I have supported numerous enterprise and institutional clients in the design of their call centers. My experience in this field includes the use of both speaker-independent and speaker-dependent speech recognition algorithms in different applications utilizing isolated speech recognition and continuous speech recognition.

17. I have worked on behalf of the Metropolitan Transportation Authority (MTA) on relay services and assistive technologies. For example, I worked on voice processing for captioning of live meeting proceedings, which are then made available for the deaf and hard of hearing. For the MTA, I implemented a system for captioning live meetings' proceedings for the deaf and hard of hearing in compliance with the Americans with Disabilities Act (ADA). The project included logistics, audio routing, voice processing, video display and captioning to generate a transcription of a speaker's voice

at a public meeting, and display of the resulting transcription on television monitors in the boardroom as well as over institutional broadcast channels and Internet feeds. In-line corrections would be generated and displayed by the transcriptionist during the proceedings. I have also worked on captioning or transcription of previously recorded information for compliance and record-keeping purposes.

18. For example, with telecommunication services and associated devices for the switching and transmission of voice and data over POTS lines, digital transmission facilities (such as T1, E1 and ISDN PRI and BRI), cellular voice and data and other wireless–based technologies (Line of Sight Microwave, Land Mobile Radio, Satellite) and IP- and VoIP-based communications, I have deployed various types of switching and routing systems (supporting numerous voice and data communications networks) at major institutional clients such as Citicorp NY, City University of New York, Bear Stearns, Societe Generale, RPCI, and Bertelsmann. For these projects, among other things, I had to understand the technical implications, shortcomings, and advantages of deploying different types of switching systems interconnecting voice and data channels, and the manner in which various communication protocols transmit voice and data and would use available bandwidth on different transmission facilities. The transmission facilities included both terrestrial (wired and wireless) as well as satellite-based communications. I also had to understand the compatibility issues with various end-user devices and central network components, both at my client's locale as well as with the telephone provider/network provider's switching systems. This encompassed both signaling and switching protocols and interfaces to support compatibility and interoperability.

19. With respect to speech recognition technology, I have worked on numerous projects that incorporated speech recognition technology in interactive voice response, dictation, and messaging systems. For example, some of my clients used speech recognition software to convert

voice messages into text that could be sent to the recipient via email or a text message (so-called unified messaging systems). For numerous clients, I deployed speech recognition technology into Interactive Voice Response (IVR) systems, improving the user experience as well as (in many instances) expediting the process flow for various real-time inquiry/response and transaction processing applications. For another of my clients, I worked on refining speech recognition protocols to be used in their telecommunications system for directory assistance. Other applications included use of speech recognition for medical transcription as well as record-keeping and compliance purposes. I have used, applied, and deployed multiple speaker-dependent and speaker-independent solutions.

20. I also have extensive experience related to the digitization of voice. For many of my projects related to call routing and switching, calls can come into the systems as digital communications or as analog POTS calls. I have frequently designed and implemented systems that can detect the digital or analog nature of the incoming voice and when analog, utilize an appropriate sample rate and codec to convert the analog voice into digital transmission. Under the same umbrella, I have designed and implemented voice messaging systems that digitally record an analog voice message in order to record or perform other digital processing on the voice signal as described above. For example, for DARPA (Defense Advanced Research Projects Agency), I performed comparisons among various voice digitization algorithms' performance in the presence of noise and jamming. As another example, for both DARPA and Defense Communications Agency (DCA), I conducted studies comparing circuit, packet and hybrid or integrated (circuit/packet) switching systems of voice and data traffic comparing them relative to cost, expandability, complexity, performance and reliability/survivability. As part of typical reliability considerations, I would deploy manual and automatic recovery processes which provided fault-tolerant and resilient voice and data networking

offering graceful degradation in the presence of errors or failures. Techniques employed encompassed error detection/retransmission, vigilant monitoring of correct operation coupled with switchover to backup, bypass and/or rerouting around failed switching and transmission facilities and reconnection.

21. As previously outlined, I have supported the deployment of several systems for generating transcriptions of previously recorded and real-time audio. For example, to support a medical dictation application, I deployed a system that used speech recognition to generate transcripts, both manual correction and auto text correction, of medical terms that were used, to correct the medical practitioner's dictation. For ADA-compliant broadcast of textual messages for the deaf and hard of hearing public at MTA meetings, I designed, selected, and deployed remote text captioning of meetings with remote in-line manual correction of errors in real-time. I also assessed captioning service accuracy and latency with the goal of improving service levels.

III. The Asserted Patents

22. The Asserted Patents are both entitled "Robust Voice Browser System and Voice Activated Device Controller," and they share a common specification. The '084 patent is a descendent of the '431 patent, with a series of intervening continuations. The patents teach several embodiments. Generally speaking, the Asserted Patents claim certain voice-controlled information retrieval systems. (Other embodiments described, but not claimed in the Asserted Patents, relate to a voice responsive device controller, like "smart home" systems in use today.)

23. The face of the patents claim ultimate priority back to the filing of a provisional application on February 4, 2000. While the conception and reduction to practice of the inventions taught and claimed in the Asserted Patents may have occurred earlier than this date, for purposes of my analysis, I view the patents through the lens of one of ordinary skill in the art in the general timeframe of late 1999 to early 2000.

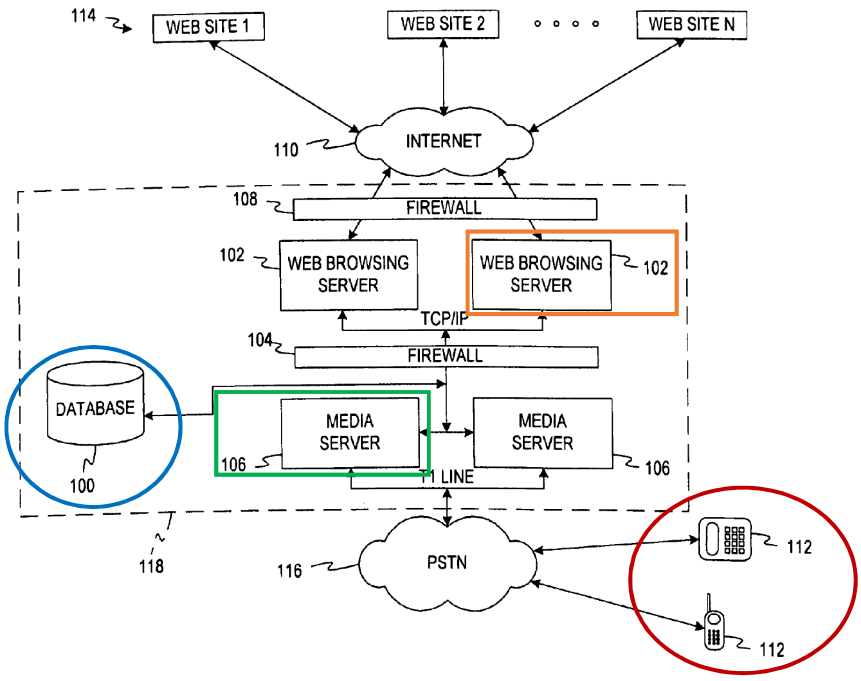
24. The specification describes, among other things, embodiments directed to a voice browsing "system for acquiring information from sources on a network, such as the Internet." *See* '431 patent at Abstract.[1] "In response to a speech command received from a user, a network interface system accesses the information source … in order to retrieve information requested by the user." *Id.* One object of an embodiment of the invention is "to allow users to gather information from web sites by using voice enabled devices," and also to provide a system that "allows the searching and retrieving of publicly available information by controlling a web browsing server using naturally spoken voice commands." *Id.* at 2:62-3:3.

25. The shared specification also describes the preferred embodiment as "allow[ing] users to access and browse web sites when they do not have access to computers with Internet access. This is accomplished by providing a voice browsing system and method that allows users to browse web sites using conversational voice commands spoken into any type of voice enabled device[.]" Further,

> These spoken commands are then converted into data messages by a speech recognition software engine running on a user interface system. These data messages are then sent to and processed by a network interface system. This network interface system then generates the proper requests that are transmitted to the desired web site over the Internet. Responses sent from the web site are received and processed by the network interface system and then converted into an audio message via a speech synthesis engine or a prerecorded audio concatenation application and finally transmitted to the user's voice enabled device.

'431 patent at 3:39-56. Figure 1 of the shared specification depicts an embodiment taught in the Asserted Patents, which is reproduced below:

[1] As the teaching of the respective specifications of the Asserted Patents is identical, I will cite only to the disclosure of the '431 patent for efficiency's sake, unless otherwise specified.



26. Some of the constituent elements in Fig. 1 are relevant to determining the claim construction analysis of the disputed terms. These include, for example, "database 100" circled in blue in the diagram above, two items "web browsing server 102" one of which is boxed in orange, and two items "media server 106" one of which is boxed in green. Also circled in red are items 112, identified in the specification as depictions of the "user's voice enabled device." '431 patent at 6:60.

27. I discuss these items in more detail below where and as necessary to inform and support my analysis. Generally, however, the specification describes "database 100" as containing

> a separate set of records for each web site accessible by the system. An example of a web site record is shown in FIG. 2. Each web site record 200 contains the rank number of the web site 202, the associated Uniform Resource Locator (URL) 204, and a command that enables the appropriate 'extraction agent' 206 that is required in order to generate proper requests sent to and to format data received from the web site.

'431 patent at 5:3-11. Here is Figure 2:

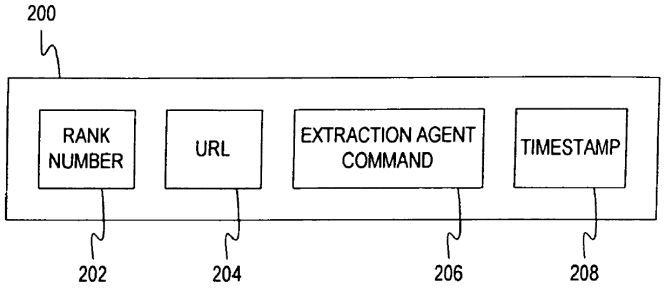


FIG. 2

28. For its part, "media server 106" operates as the "user interface system[]." '431 patent at 5:61-62. In particular, in the preferred embodiment, the media servers 106 contain, among other items, "a speech recognition engine 300, a speech synthesis engine 302, [and] an Interactive Voice Response (IVR) application 304, a call processing system 306, and telephony and voice hardware 208[.]" *Id.* at 63-66. The specification further teaches that the "speech recognition function [of the preferred embodiment] is performed by the speech recognition engine 300 that converts voice commands received from the user's voice enabled device 112 … into data messages." *Id.* at 6:5-8.

29. The specification identifies a preferred speech recognition engine as developed by Nuance Communications, and also teaches that the speech recognition engine 300 of the preferred embodiment uses "natural speech recognition grammars (i.e., what a user can say that will be recognized by the speech recognition engine)." *Id.* at 6:16-24.

30. Additionally, the media servers 106 of the preferred embodiment also "contain a speech synthesis engine 302 that converts the data retrieved by the web browsing servers 102 into audio messags [sic] that are transmitted to the user's voice enabled device 112." '431 patent at 6:57-60.

31. The "voice enabled device 112" of the preferred embodiment is described as any type of voice enabled device such as, for example, "wireline or wireless telephone, IP phone, wireless PDA, or other wireless device." '431 patent at 2:62-65, 3:16-19, 3:39-46, 23:50-52. Also of note is

Claim 4 of both Asserted Patents. Each claim 4 claims a system wherein the "voice enabled device is a standard telephone, an IP telephone, a cellular phone, a PDA, a personal computer, a DVD player, a television or other video display device, a CD player, a MP3 player, or any other device capable of transmitting said audio message." *Id.* at 20:45-50.

32. Moreover, the specification describes "web browsing server 102" as providing "access to any computer network such as the Internet 110 … [and the] web browsing servers receive responses from web sites and extract the data requested by the user." '431 patent at 6:65-7:4. Once the web browsing server retrieves the requested information, it "is forwarded to the media server 106" for conversion into an audio message to be delivered to the user's voice enabled device. *Id.* at 32-37.

IV. The Relevant Technology

33. In addition to the teachings of the Asserted Patents, it may be helpful to offer some high-level background regarding the relevant technology. As noted, the patents are directed to voice-controlled information retrieval systems. Several aspects of the relevant technology are key to understanding the claimed inventions. At a high level, these aspects can be grouped into two main categories: (1) voice control, and (2) information retrieval. I address each category in turn.

A. Voice Control

34. Generally speaking, voice control is the ability of a system to recognize and translate spoken language requests or commands into text or data, and to take an action in response to such spoken language commands. This requires speech recognition technology.

35. Speech recognition is the ability of a system to recognize and translate spoken language into text. Specifically, speech recognition answers the question: "What did he/she say?" Some common examples of speech recognition applications include voice user interfaces, such as voice dialing (e.g., "call mom"), appliance control, data entry (e.g., providing a credit card number to

a call center), and dictation (e.g., used in transcription, captioning, and messaging in conjunction with word processing, text editing or email systems).

36. Automatic Speech recognition (ASR) systems are characterized by their capabilities, primarily the language, vocabulary size (number of words that can be recognized in the language), the ability to recognize continuous speech or merely isolated words, and whether the system is designed for use by one speaker or can be used by many speakers. The principal performance criterion is word accuracy (the percentage of words that are correctly recognized). While the accuracy depends primarily on the algorithm used, there are other dependencies that will impact ASR performance, including background acoustic noise, channel noise (if speech is transmitted via telecommunications networks), microphone quality and placement and other subtleties such as speaker dialect/accents, articulation, stress and nasality (e.g., due to colds). Other metrics include the computing power required to execute the algorithm (in real time), associated processing latency to generate the text (from the uttered speech), and the corresponding cost of the ASR system.

37. Over nearly sixty years of progress, ASR has made dramatic improvements in performance. ASR has evolved from supporting applications which could only perform isolated word recognition for a very small vocabulary for a single speaker, to recognizing continuous natural language speech for a very large vocabulary on behalf of many different speakers. The technical advances have been made possible by breakthroughs in the algorithms used to recognize speech driven by the efforts of many researchers and practitioners, often adopting an interdisciplinary approach drawing upon expertise in diverse fields including electrical engineering (e.g., digital signal processing), mathematics (e.g., probability and statistics), computer science (e.g., artificial intelligence), acoustics, linguistics, phonetics, physiology, and even psychology. Bringing the theory to practice, technological advances in raw increases in computing power, specialty processors (such

as DSPs), and distributing the recognition process have enabled the real-world implementation of these solutions.

38. In what follows, I briefly discuss some of the features and capabilities of speech recognition algorithm more or less representative of and contemporaneous with the filing of the Asserted Patents. In general, ASR systems are comprised of an acoustic and a linguistic model. The acoustic model typically exploits inherent characteristics of human speech representing the important properties of speech (including voicing, pitch, and the shape of the vocal tract). Acoustic feature extraction can also be used to qualify and distinguish one speaker's inherent characteristics from another and, for purposes of speaker identification and/or verification in some applications, may be stored as part of an overall user profile (e.g., feature vectors). The acoustic model also typically deals with phonemes (i.e., the basic sound units of speech that comprise syllables and in turn words). A "phoneme" is a unit of sound in a language that cannot be analyzed into smaller linear units and that can distinguish syllables within a word allowing the ability to discern one word from another. I attach the entry for "phoneme" from the Oxford English Dictionary confirming this definition. *See* **Exhibit B**.

39. In addition to stringing together syllables to form words, the linguistic model uses syntactic (and in some instances semantic) dependencies embedded in the language (also influenced in part by the vocabulary). These are directly taken into account (e.g., what words are likely to follow another word). Some algorithms will use groups of phones (the actual sound of the phoneme) to determine what sound was actually uttered. More sophisticated models will also examine word clusters (looking forward and looking backward) in formulating hypotheses regarding the most probable word or word combinations; this improves recognition accuracy by establishing context.

40. Some early speech recognition systems utilized what is now known as "isolated" speech recognition. In other words, the user needed to inject an artificial pause between each word. This was obviously unnatural and slow. Later, "continuous" speech recognition was achieved, which removed the requirement to inject the unnatural pause between words and solved the challenges posed by lack of word separation, co-articulation, and speaking at variable rates. Significant advances that influenced the acoustic and linguistic models and were a fixture (even to the present day in some algorithms) are the Hidden Markov Model (HMM) which is a statistical model to interpret a sequence of sounds and formed the cornerstone of many modern ASR systems, and the n-gram language model. An n-gram is a sequence of n items, typically words, and the algorithm uses the probability that a given word was uttered as a function of the previous n-1 words uttered in a sentence.

i. Speaker-Dependent versus Speaker-Independent speech recognition

41. An important aspect of relevant speech recognition technology to appreciate for purposes of claim construction in this case is the difference between "speaker-dependent" and "speaker-independent" speech recognition systems.

42. Speaker dependent systems are designed for use by one individual, an individual to which that system must be trained (often extensively). A traditional benefit of these systems is that they support larger vocabularies and tend to be very accurate—but they are typically optimized for use by only one person, and only after that single user invests significant time training the system for use. The typical application where such algorithms are used is dictation, where the transcription system is optimized for a particular speaker's exclusive use. In these applications the vocabulary is large and there may also be the opportunity for the user to "correct" the transcription allowing the ASR to learn from its mistakes.

43. On the other hand, speaker independent systems are designed for applications which support many speakers—but do not have the luxury of an extensive amount of time for training the system by a given user (i.e., the interaction must be prompt). An obvious benefit of these systems is that they are designed to be used "out of the box" or upon first encounter by any individual. One example of a classic speaker independent system is that of an automated call center, which businesses use to great effect to service a variety of callers. Significant success has been achieved in call center applications using interactive voice response, with users being prompted by the IVR to speak their credit card numbers, or speak a keyword from a menu of options, as opposed to entering 1, 2 or 3 from their telephone's keypad.

44. Another example of speaker independent speech recognition is the use of keyword spotting in call centers or a 911 Public Service Answering Point (PSAP). In these applications, the ASR can constantly "listen" to a call-in-progress—not necessarily converting the speech into text but monitoring the conversation. Upon utterance of a volatile word (e.g., "kill" or "bomb"), the ASR will alert ancillary systems that the call should be recorded, and/or that a supervisor should be automatically bridged into the call. In general, though, speaker independent ASR systems are less accurate than the extensively trained speaker dependent counterparts when applied to the same vocabulary.

45. To address the lower accuracy of speaker independent systems, practitioners in the field began to experiment in the 1990s with adapting or modifying the systems using data collected from one or more users to better and more accurately function as intended.

ii. Training and Adaptation

46. Relatedly, there is an important distinction between "training" in the context of speaker dependent systems on the one hand, and "adaptation" in the context of speaker independent

systems on the other hand. At a high level, the principal difference is that training is necessary before any reasonably accurate use of a speaker dependent system is possible. By contrast, adaptation may be used (but is not necessary) to fine-tune or otherwise enhance the accuracy of a speaker independent system.

47. This distinction between training and adaptation in the speech recognition context is reflected in academic writing in the field in the 1990s. For example, Xuedong Huang, then a researcher at Carnegie Mellon, authored a paper called "A Study on Speaker-Adaptive Speech Recognition" in which he observed that "[s]peaker-independent systems are definitely desirable in many applications where speaker-specific data do not exist. On the other hand, if speaker-dependent data are available, the system could be adapted to a specific speaker to further reduce the error rate." This same author posited that a logical compromise for a practical system is to start with a speaker-independent system, and then adapt the system over time to each individual user. I attach a copy of this paper as **Exhibit C** (PARUS_00008718-PARUS_00008723).

48. Others in the field over time also recognized that speaker-independent systems could utilize information obtained from users to enhance accuracy, but distinguished this approach from the type of initialization training required for speaker dependent systems. As just one example, in 1995 Leggetter and Woodland authored a paper titled "Maximum likelihood linear regression for speaker adaptation" in Computer Speech and Language. I attach this paper at **Exhibit D** (PARUS_00008724-PARUS_00008738). They noted, among other things, that "it is desirable to use a small amount of [a] new speaker's speech (adaptation data) to 'tune' the SI [speaker independent] models to the new speaker[.]" In addition, Richard Cox and his team presented a paper called "Speech and Language Processing for Next-Millennium Communications Services" in August 2000, noting the benefits of

adaptation in the broader context of natural language understanding. I attach this paper as **Exhibit E** (PARUS_00008685-PARUS_00008708).

B. Information Retrieval

49. The second main area of relevant technology to the claim construction analysis in this case is that of information retrieval, and specifically the retrieval of information via a network. One may retrieve information over any network, such as one that is local, such as information stored on a private server. This information can be accessed by locally-attached PCs or terminals, users on a wired or wireless local area network, or even remote users through a wide area network, all with suitable access credentials. These networks are often referred to as "intranets." More commonly used is the kind of information contained within the World-Wide Web (WWW) (which is the collection of websites and other web resources containing information) retrievable over the Internet. The Internet is a publicly accessible network of interconnected computer networks that are resident outside of one's local network. The WWW server is typically hosted by a third party or any public Internet segment operated by a private organization, accessible by connected computer using Internet protocols.

50. One way to access information on the Internet is using a typical web browser, such as Internet Explorer (IE) or Chrome (or, at the time of the patents, IE or Netscape Navigator). This allows a user to obtain information viewable on, or accessible through, a web page. Typically, this kind of information is coded in what is called Hypertext Markup Language, or HTML, a frequently used "markup language" for hypermedia designed to be displayed in a web browser. Hypermedia is a nonlinear medium of information that includes graphics, audio, video, plain text and hyperlinks. In addition, Hypertext Transfer Protocol, HTTP, is an application-layer protocol for transmitting hypermedia documents like HTML. HTTP was designed for communication between web browsers and web servers, but it can also be used for other purposes.

51. To automate manual data extraction on websites, web crawlers (also referred to as "spiders") automatically browse the Internet to index and search for content. In this way, content accessible via the Internet is available not only to an individual browsing web pages using a web browser, but through an automated spider. Once a suitable web site and webpage of interest is accessed by the spider, a web data extraction is performed (sometimes called "scraping,") to access specific content. A web scraper is a specialized tool that can access the detailed contents available at, or through, a web site. It extracts information from specific pages by parsing the contents in the markup language and extracting the data, and then converting it into a format that can be used (in the raw by the monitoring application) or potentially stored in a database for backend information accessible by other applications.

V. The Appropriate Definition of a Person Having Ordinary Skill in the Art

52. Based on my review of the teachings of the Asserted Patents, in my opinion a person of ordinary skill within the field of the Asserted Patents, which is voice-controlled information retrieval systems, around the late 1999 to early 2000 time period would have an electrical or computer engineering background and practical experience in the field. Such a person would have a Bachelor's degree in electrical or computer engineering, or in a related field, and at least two years of work experience relating to web-based information retrieval systems, speech recognition and interactive voice response systems, or related systems.

53. In view of this definition, I consider myself to be at least one of ordinary skill in that art at the time of the invention. As already discussed, I obtained a Master's Degree in Electrical Engineering in 1975 and, by the time of the invention, I had been working in the relevant technology space for approximately 30 years, if not more.

VI. Standards for Claim Construction

54. I am not an attorney, and offer no legal opinions. Nonetheless, I have a general understanding based on my discussions and work with patent attorneys over the years regarding certain legal standards that are relevant to assessing the appropriate meaning of disputed patent claim terms. I describe my understanding of these standards below.

55. I understand that a patent includes at least claims and a specification. The specification includes a description of the invention and drawings, and may include a description of the preferred embodiment of the invention. The claims define the scope of the patented invention.

56. I understand that the purpose of claim construction is to determine the meaning and scope of the asserted claims. When interpreting the claims, I understand that the ordinary meaning of the language within the claims should generally be followed.

57. If the claim language is not clear on its face, I understand that the rest of the intrinsic record, including the patent specification and file history, can be consulted to resolve any lack of clarity in the claim terms themselves. I further understand that the claims cannot be interpreted to require a particular feature of an embodiment of the invention described in the specification, even if it is a preferred embodiment, unless the claims expressly recite those features.

58. It is my understanding that the claim terms should be construed to have the meaning that a person of ordinary skill in the art at the time of the invention would employ when reading the claims. However, such a reading must be made within the context of the claims and the entire intrinsic record. Thus, for example, if the patentee decides to clearly and unequivocally define a term in an unconventional manner, that definition should be honored when reading the patent. I also understand that the manner in which a claim term is used within the patent specification should be given significant weight when determining claim meaning.

59. Further, extrinsic evidence (evidence beyond the claims, the specification, and the file history) may be used to enhance understanding of the patented technology, but not for the purpose of varying or contradicting claim terms. It is my understanding that extrinsic evidence is consulted only when claim language remains genuinely ambiguous after consideration of the intrinsic evidence.

VII. The Disputed Claim Terms

60. I have reviewed in detail the claims, shared specification, and potentially relevant portions of the respective file histories of the Asserted Patents. I have also reviewed some of the extrinsic evidence disclosed by the parties on May 5, 2020.

61. I am aware of the claim terms of the Asserted Patents that Parus and the Defendants intend to put to the Court for construction, and the competing proposed constructions. At a high level, I agree with Parus's position with regard to each disputed claim term. Below I describe why, and offer my opinions as one skilled in the art regarding each such claim term in turn.

62. I am also aware that Defendants contend that certain claim terms are indefinite and/or governed by § 112, ¶ 6. With respect to these contentions, I understand that Defendants have not yet articulated any substantive reason why they believe such claim terms are indefinite and/or means-plus-function. I reserve the right to offer additional opinions once Defendants and/or their expert(s) offer more than the allegation that certain claims are indefinite and/or means-plus-function.

A. "voice enabled device"

63. I agree that the appropriate articulation of the meaning of the claim term "voice enabled device" is as Parus proposes: "a wired or wireless voice communication device associated with audio input (e.g., a microphone) and audio output (e.g., a speaker) capabilities." This comports with my understanding of the plain and ordinary meaning of the claim term to one skilled in the art at the time of invention.

64. On the other hand, in my opinion Defendants' proposed construction, "wireline or wireless telephone, IP phone, wireless PDA, or other wireless device," is too narrow. For example, it would seem to exclude devices that may communicate with a network via WiFi but are plugged into a wall outlet, i.e., does not run only on a battery. Such a restriction is inconsistent with the way the claim term is used in the Asserted Patents.

B. "speaker-independent speech recognition device"

65. Parus's proposed articulation of the plain and ordinary meaning of the claim term "speaker-independent speech recognition device" is also accurate, in my opinion. As the clear teaching of the shared specification discloses, such a device need not be trained in order to be used. In particular, the specification states: the contemplated speaker-independent speech recognition device can "recognize[] naturally spoken voice commands and is speaker independent; it **does not *have* to be trained** to recognize the voice patterns of each individual user." '431 Patent at 4:38-43 (emphases added). As one skilled in the art at the time of invention, this comports with my understanding of the term as it is used in the Asserted Patents.

66. In general, speaker dependent systems are typically used for single speaker large vocabulary applications (such as dictation), and are optimized for one speaker's voice requiring extensive training. By contrast, speaker *in*dependent systems are typically used in interactive voice recognition ("IVR") systems where extensive training is impractical, or in systems designed to support large numbers of speakers and therefore typically do not *require* extensive training. Unlike Defendants' competing proposals, Parus's articulation of "speaker-independent speech recognition device" does not preclude altogether any kind of adaptation or other tailoring to an individual user or users. For example, the specification teaches that in one embodiment the system retains user profile information. '431 Patent at 5:55-56 ("database 100 may contain customer profile information"). In

my opinion, this suggests to one skilled in the art that such information could include data used for just that purpose, much like the type of "tuning" data (e.g., the speaker's acoustic features extracted in the acoustic model of the ASR) and contemplated by practitioners in the 1990s discussed above.

67. It is my opinion that Defendants' proposed constructions are insufficient to capture the proper scope of the term "speaker-independent speech recognition device." This is because each definition expressly excludes any kind of tuning or adaptation to particular users, and is inconsistent with the meaning of the term as explained in the intrinsic evidence.

C. "recognition grammar"

68. With respect to the claim term "recognition grammar," in my opinion Parus's proposal is the only one that aligns with my understanding of the term as one skilled in the art. It is also the proposal that accurately reflects the teachings of the shared specification. In particular, the specification clearly teaches that the "recognition grammar" contemplated by the invention is "what a user can say that will be recognized by the speech recognition engine." '431 Patent at 6:21-24. This is exactly Parus's proposal.

69. In addition, in my opinion the two different proposals offered by Defendants are not consistent with how one skilled in the art would understand the term. These proposals each require that the claimed recognition grammar must consist of "predefined" words, phrases, and/or rules. It is unclear to me what these additional limitations are intended to accomplish for purposes of clarifying the claim term.

70. In any event, I don't believe that these limitations are supported by the intrinsic record, as I am unable to discern anything in the intrinsic record suggesting that the claim term "recognition grammar" must include Defendants' "predefined" words, phrases, and/or rules. The vagueness in this aspect of Defendants' proposed constructions only compounds the problem with their proposals. For

example, when must the words, phrases, and/or rules be predefined in an infringing system? Can those words, phrases, and/or rules change? It is unclear from Defendants' constructions.

D. "web site"

71. In my opinion, the claim term "web site" needs no particularized or specific meaning. One skilled in the art would readily understand this term as used in the Asserted Patents. Defendants' proposal is thus unnecessary, but in my opinion it is also incorrect. As I read it, Defendants' proposal seeks to limit the term to only web pages, plural, that can be accessed by a user using a web browser like Internet Explorer. That does not capture the proper scope of the term "web site" as understood by one skilled in the art.

72. For example, this would exclude web sites that consist of only one web page. It would also exclude other information that is resident on and/or accessible through a web site, but that might not be viewable to a user visiting the site using a web browser. Such information might include, for example, a white paper in PDF form downloadable from the web site, or the HTML underlying the viewable web pages of that web site. In my opinion, these examples would be well within the scope of the term "web site" as understood by one skilled in the art, and should therefore not be excluded from the definition as Defendants contend. This is consistent with the shared specification, which discusses a web browsing server 102 that automatically retrieves web-accessible information from the web-enabled sources in response to voice commands from the system's user. *See, e.g.*, '431 Patent at FIG 1, FIG 4, and 4:54-57 ("In the first preferred embodiment, the network interface system is referred to as a web browsing server.")

E. "select the corresponding recognition grammar upon receiving [said/the] speech command"

73. In my opinion, the claim term "select the corresponding recognition grammar upon receiving [said/the] speech command" needs no particularized or specific meaning. One skilled in the art would readily understand this term as used in the Asserted Patents. The proposal offered by the Android Defendants and Amazon ("upon receiving [said/the] speech command, select the recognition grammar for the corresponding category of information") is not required, not least because it appears to merely rearrange the words of the claim term itself without apparent justification. In my opinion, it also appears to limit the claim through the addition of "the corresponding category of information" limitation. I do not believe that this limitation is required by anything in the intrinsic record. Adopting the proposal of all Defendants except for Apple would not, in my opinion, capture the proper scope and meaning of the term "select the corresponding recognition grammar upon receiving [said/the] speech command" as understood by one skilled in the art.

F. "access at least one of [said/the] plurality of web sites identified by [said/the] instruction set to obtain [said/the] information to be retrieved"

74. In my opinion, the term "access at least one of [said/the] plurality of web sites identified by [said/the] instruction set to obtain [said/the] information to be retrieved" should be given its plain and ordinary meaning. It would be readily understood by one skilled in the art reading the claims in view of the teaching of the specification. As a result, I do not see any need to give the term a specific articulation or construction. And, even if I did, I would disagree that Defendants' proposed construction is proper. In particular, as one skilled in the art, I do not interpret the term "access" to be "search," as the Android Defendants propose. Even from a high-level, plain English perspective, I don't see how "access" can possibly be understood as "search." I simply don't see the basis for such a change.

Signed under the pains and penalty of perjury, this Fifth day of June, 2020.

Benedict Occhiogrosso

CERTIFICATE OF SERVICE

The undersigned hereby certifies that a true and correct copy of the foregoing document has been served on June 5, 2020 by email to all counsel of record.

/s/ *Andrew H. DeVoogd*
Andrew H. DeVoogd